

# Weekly Report

2012.09.03-2012.09.09

黄芯芯

## 本周工作：

1. 淘宝标签项目：
  - 1) 通过 Python 脚本从淘宝 TCIF 给出的接口中爬了 10000 个用户的性别与年龄数据，耗时需要 3 个小时左右。以后爬四十万个用户数据需要耗时更多。
  - 2) 调整代码结构，加快 KDE 的计算。现在每次计算需要几秒的时间，相对此前几分钟的计算耗时有很大提升。
  - 3) 参考 colorbrewer: <http://colorbrewer2.org/>，调整了配色方案，如下图所示。将颜色划分为冷色与暖色两种，KDE 结果低于中间值的取蓝色，高于中间值的取红色。然后蓝色和红色的浓度与 KDE 结果的大小成正比，即 KDE 结果越大浓度越高，表示在该区域越密集。

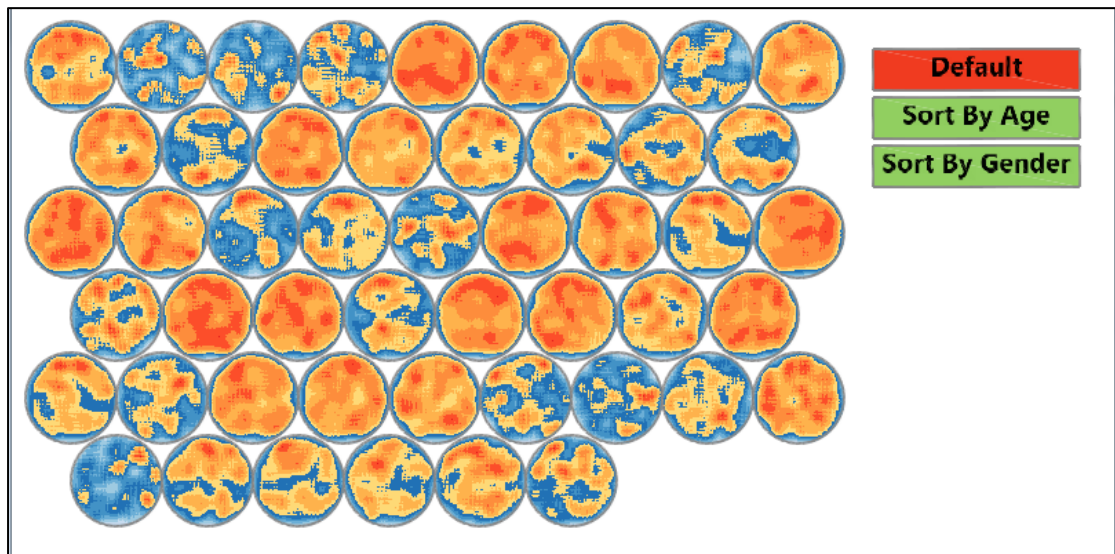


图 1：按用户 ID 排序结果

- 4) 界面上加入了将用户按照年龄或性别进行排序的选项，结果分别如图 2、图 3 所示。同时，鼠标停留在某个类目上时，该类目名称会显示。

从图 3 按年龄排序的结果上看，首先可以注意到图中红色方框的三个类目有相似的 PATTERN，查看这三个类目，分别是“腾讯 QQ 专区”、“游戏装备/游戏币”和“网络点卡”，果真是存在关联。将年龄进行排序后，对于如何能够将年龄的信息很好的展示出来还没有想到很好的办法。

图 4 的结果是将用户按照女-男的方式进行排序的结果，结果呈现出比较明显的男女分化。

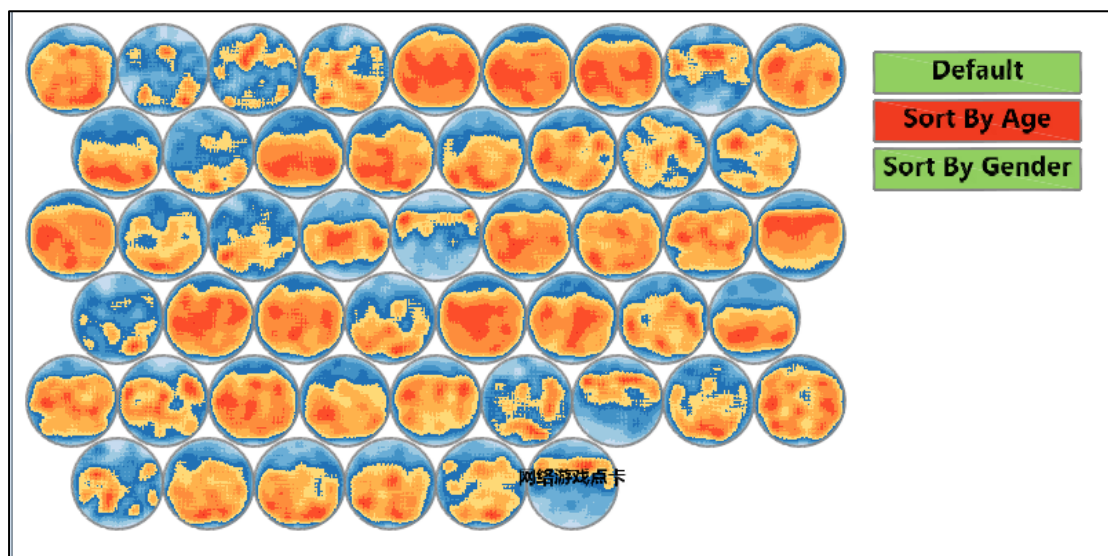


图 2：按年龄将用户数据排序

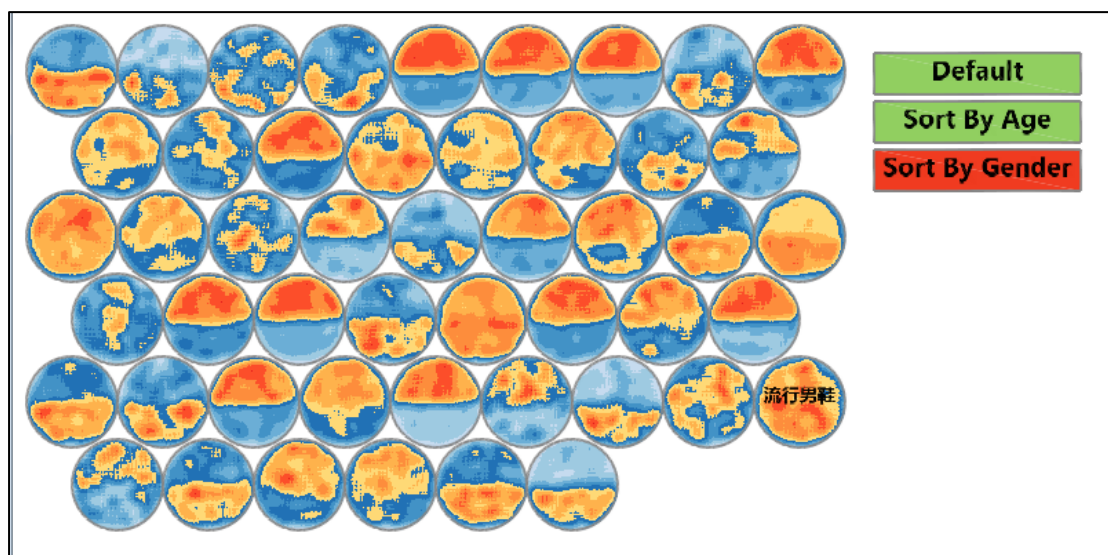


图 3：按性别将用户数据进行排序

## 5) 遗留问题：

- 如何更好地将这 50 个类目进行有意义的布局？
- 如何单独展示男女性别用户的数据？
- 如何将年龄信息呈现出来？

## 2. DATAV.JS 组件库：

这周三开了一次关于组件库第二期规划的会议。会议安排是接下来一个月时间完成组件库的一些代码重构、开发基础组建、接入“在云端”项目等目标。

（这周实际工作时间只有三天，后面几天因为新生大会、体检的事情天天往玉泉跑。）

### 下周计划：

1. 思考项目遗留问题。
2. 开始秋学期的课程。
3. 秋学期出去上课的时间，只有两天的时间可以去实习，关于这个要和玄澄沟通一下。